

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 6, 2001, 18:42:29 ; Search time 3825.23 Seconds
(without alignments)
8767.220 Million cell updates/sec

Title: US-09-494-297-1
Perfect score: 2274
Sequence: 1 atgataaaacaaagtcttc.....ggataagaaacatgactag 2274

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1283235 seqs, 7373929652 residues
Total number of hits satisfying chosen parameters: 2566470

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 08
Maximum Match 1008
Listing first 45 summaries

Database :

GenEmbl:*

- 1: gb_ba1:*
- 2: gb_ba2:*
- 3: gb_ba3:*
- 4: gb_in1:*
- 5: gb_in2:*
- 6: gb_in3:*
- 7: gb_cm:*
- 8: gb_ov:*
- 9: gb_pat1:*
- 10: gb_pat2:*
- 11: gb_ph:*
- 12: gb_pl1:*
- 13: gb_pl2:*
- 14: gb_pl3:*
- 15: gb_pl4:*
- 16: em_ba1:*
- 17: em_ba2:*
- 18: em_fun:*
- 19: em_hcgo_hum:*
- 20: em_hcgo_inv:*
- 21: em_hcgo_rod:*
- 22: em_hcgo_hum1:*
- 23: em_hcgo_hum2:*
- 24: em_hcgo_hum3:*
- 25: em_hcgo_hum4:*
- 26: em_hcgo_hum5:*
- 27: em_hcgo_hum6:*
- 28: em_hcgo_hum7:*
- 29: em_hcgo_hum8:*
- 30: em_hcgo_inv1:*
- 31: em_hcgo_inv2:*
- 32: em_hcgo_other:*
- 33: em_hcgo_rod:*
- 34: em_hum1:*
- 35: em_hum2:*
- 36: em_hum3:*
- 37: em_hum4:*
- 38: em_hum5:*
- 39: em_hum6:*
- 40: em_hum7:*
- 41: em_in:*
- 42: em_cm:*
- 43: em_or:*

44: em_ov:*

45: em_pat:*

46: em_ph:*

47: em_pl:*

48: em_ro:*

49: em_sts:*

50: em_sy:*

51: em_un:*

52: em_v1:*

53: gb_sts1:*

54: gb_sts2:*

55: gb_sts3:*

56: gb_sy:*

57: gb_un:*

58: gb_v11:*

59: gb_v12:*

60: gb_hc1:*

61: gb_hc2:*

62: gb_hc3:*

63: gb_hc4:*

64: gb_hc5:*

65: gb_hc6:*

66: gb_hc7:*

67: gb_hc8:*

68: gb_hc9:*

69: gb_hc10:*

70: gb_hc11:*

71: gb_hc12:*

72: gb_hc13:*

73: gb_hc14:*

74: gb_hc15:*

75: gb_hc16:*

76: gb_hc17:*

77: gb_hc18:*

78: gb_hc19:*

79: gb_hc20:*

80: gb_hc21:*

81: gb_hc22:*

82: gb_hc23:*

83: gb_hc24:*

84: gb_hc25:*

85: gb_pr1:*

86: gb_pr2:*

87: gb_pr3:*

88: gb_pr4:*

89: gb_pr5:*

90: gb_pr6:*

91: gb_pr7:*

92: gb_pr8:*

93: gb_pr9:*

94: gb_rol:*

95: gb_rol2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
c 1	720.8	31.7	10826	3	SPU49397
2	215.8	9.5	2845	3	STREPTINF
3	180.2	7.9	2263	3	SPSFB
4	167.6	7.4	2066	2	GU31115
5	150.2	6.6	2801	1	AF009908
6	97	4.3	321	1	AF009914
7	92.8	4.1	313	1	AF009913
8	72	3.2	348	1	AF009920
9	69.8	3.1	149789	60	AC007926
10	68.4	3.0	309	1	AF009915
11	68.2	3.0	278	1	AF009917

```

12 68.2 3.0 195969 76 AC074354
13 67.6 3.0 281 1 AF009909 Streptoco
14 67.6 3.0 282 1 AF009912 Streptoco
15 66.8 2.9 310 1 AF009911 Streptoco
16 64.2 2.8 326 1 AF009910 Streptoco
17 64.2 2.8 2442 5 AF103869 Streptoco
18 61.6 2.7 309 1 AF009918 Streptoco
19 61.4 2.7 255 1 AF009916 Streptoco
20 60.2 2.6 6941 6 ENMHGAX
21 59.4 2.6 256172 60 AC005139
22 59.4 2.6 310779 60 AC005140
23 59 2.6 83110 84 PFMAL13PB
24 58.8 2.6 4102 5 AF270648
25 58.6 2.6 308 1 AF009919
26 58.6 2.6 879 53 CNS01JRG
27 58.6 2.6 1007 53 CNS06X9S
28 57.4 2.5 131184 92 HSH2137
29 57.4 2.5 340000 36 HSH21C017
30 57.4 2.5 340000 90 AP001672
31 56.6 2.5 12240 4 AE001424
32 56.6 2.5 232392 58 AF250284
33 56.4 2.5 321003 84 PFMALAP3
34 56 2.5 17470 4 AE001402
35 55.8 2.5 1137 53 CNS071NS
36 55.4 2.4 694 6 S52459
37 55.2 2.4 897 54 CNS07ABZ
38 55.2 2.4 3711 5 AF247634
39 54.6 2.4 759 53 CNS07E8R
40 54.4 2.4 964 54 CNS06GXV
41 54.4 2.4 174832 83 CNS01RGQ
42 53.8 2.4 9432 4 AE001375
43 53.6 2.4 110000 84 PFMAL13P2_0
44 53.6 2.4 156060 60 AC004153
45 53.6 2.4 204652 84 PFMAL13P6

```

ALIGNMENTS

```

RESULT 1
LOCUS SP049397/c 10826 bp DNA BCT 19-DEC-1998
DEFINITION Streptococcus pyogenes Msmr (msmr) gene, partial cds; Lepa (lepa),
partial cds; and Nra (nra) genes, complete cds; Ssba (ssba) gene,
partial cds; and unknown genes.
ACCESSION U49397
VERSION U49397.1 GI:4028947
KEYWORDS
SOURCE Streptococcus pyogenes.
ORGANISM Streptococcus pyogenes
Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
Streptococcus.
REFERENCE 1 (bases 1 to 10826)
AUTHORS Podbielski,A., Woloschik,M., Leonard,B.A.B. and Schmidt,K.H.
TITLE Characterization of nra, a global negative regulator gene in group
A streptococci
JOURNAL Mol. Microbiol. 31 (1999) In press
REFERENCE 2 (bases 1 to 10826)
AUTHORS Podbielski,A. and Woloschik,M.
TITLE Direct Submission
JOURNAL Submitted (19-FEB-1996) RWTH Aachen, Institute of Medical
Microbiology, Pauwelsstrasse 30, Aachen, NRW, Federal Republic of
Germany, 52074
FEATURES
Source
1..10826
/organism="Streptococcus pyogenes"
/strain="CS101"
/db_xref="taxon:1314"
<1..630
/gene="msmr"
<1..630
/gene="msmr"
/function="multiple sugar metabolism regulator"

```

```

repeat_region
CDS
/codon_start=1
/transl_table=11
/product="Msmr"
/protein_id="AAC97150.1"
/db_xref="GI:4033713"
/transl_table="DLVKLGNPQLKOEINRIPSSSTSSISALRAEKNTLYVLR
LLEFSEVENTDAKHYSLVKYMALNEASDLKYLRCALITHFSSSLNKSISDK
ROMNSVLHYVDSHYSLKYSKDIARKLYVSESHRSYVKYSNSLOHYLSTKIKE
AQLLKRGPVGEVAKSLYFYDTTHFHRIKFKYTGSSKDYLAQRONI"
697..730
/feature="putative transcription terminator"
/rpl_type="inverted"
complement(790..1377)
/feature="ORF 2"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAC97151.1"
/db_xref="GI:4033714"
/transl_table="MRKYMKLFSVMMMLTLAFOQTVLADSTVQTSISEVNLERA
GSPFPESTALESIDAMKTIETITSGSKASPSPLNTGVGYTVRYVOKSPNDKO
ADTVYDVLYVTVYVDEGDTLAKVYSRRAGDEKSAITPKRRLVKKPIPPQDLPKT
PLPLAGEVKSLSGITSYVLGDLVLLVYKRLSL"
complement(1394..3169)
/feature="similar to electron transfer flavoprotein 1 of
Methylobacillus electron; ORF 3"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAC97152.1"
/db_xref="GI:4033715"
/transl_table="MKKNKLALATLALATLALGMSMSONIKAEAGVIGDSTLVYKKT
PSTYDDNVLMPKADSEFEVADNAKGTGDLIKGVIDGLENTKIRYNSDKI
TAKEKSVNPEEFANVKEPGVGYRVAVANGKAGITDQSOQMTDVTYVYKKEGFE
VKYIYSTERGQSEKKPVLPKNSPTTSLKIEQVYGNNGEHRLLSFLLLPNCFE
KGQVNIIDGSEFKRVVIGEYSFTLKRGSTVLSQLEVGLETKLDEEDVKGDKITS
ATLKDEQSSYELGDKDHTKDSADEIVYTNKRDQVPGVAGTAPAVLSIYVAGG
VYIYTRKRAQVYSQMMMTIYOVYNKAIDTLLIFCLVYFLAEGFLMDSYHLVQGA
DASNFKKFRTAOQCPFEDLLANEDVIGMLIPGTHIDYPLVQGTMLLEYINKAVD
SVAMSGSLFLIDRNHNDFTDVSLSLYGHMAGNMGELPKPKKFNKKNKAIET
KERKLLYTVIRACLTDAFDQVFPNMTITNDQQRVLVDYISKRSKPKVKKLKHHT
KEVAFSTCENSTDNRYIVGVGTQF"
complement(3191..3712)
/feature="Lepa"
/gene="lepa"
complement(3191..3712)
/feature="Lepa"
/feature="ORF 4; similar to Staphylococcus aureus signal
peptidase 1"
/codon_start=1
/transl_table=11
/product="Lepa"
/protein_id="AAC97149.1"
/db_xref="GI:4033712"
/transl_table="MTNYLNRINENSLKRAPIRLVLYKISITGFLYILFYQVFGMIIT
NTNDSPALSGDGLYLRADRSIINDVYEVNDTLKVGSLAAQAGDEVNFTQEGG
LLINHPPEKEVEPVLYPHSSGPNFPYVPGTVFILLDYREERLDSRYGALPTNOI
KGISTLRLVRGCI"
complement(3705..6180)
/feature="cpa"
/gene="cpa"
complement(3705..5933)
/feature="cpa"
/feature="collagen-binding protein of group A Streptococci"
/codon_start=1
/transl_table=11
/product="Cpa"
/protein_id="AAC97148.1"
/db_xref="GI:4033711"
/transl_table="MOKRDKTNYGSAANKRRQTTIGLKFVLFTEVALIGVGSIRAF
GAEROSVNPROSITODYPWYGYDSKPYSPYKTYHNKLVNLESGKDOYACENL
TKHPSKSDSVRSQYKKLLEGNTENFYLAKPKRLEDQLOQNLIRILYNYVPAWNG
IMKGIIDLNALIVTONAIWYDSAOINPDESFKTEARSNGINDOOLGIMKRALBLID
PNLGYSKSKTPSGRNLWVFPESHDKPFCNLLSAEYVPTPPKPGEEPATKEKTSVIL
RKYAEGSDKLLEGATLKLKSQIEGSGFQCKDFOKNSIGTVELPNTGYTLTFTSSPDGY

```


Db 5246 CCATCAGTTATCGTTAAATGATTGAAATCTCATGATAAGCT----- 5200
 QY 781 ggtatccaaaactcttgatggtggtttagttctactaaaccacactccagagac 840
 Db 5201 ---TTCGAAAATCTTTGAGTGTGAGTATGTCGAGTACTCCCGAAAACCGAGAGA- 5144
 QY 841 ccacaaatgctccaaatccaaactccaaagctcagtaacttaagaaatgagata 900
 Db 5145 ----AAGGCTCGCGGCTAAACGTAAGAAAACATCATCATATACAGAAAATATCGGAA 5091
 QY 901 ggtatctacttaaatgtctgaaggtgcaacatcagttgacagggagataacgtgaat 960
 Db 5090 GGTG---ACCTTAATCTTCTAGAGGAGCAACCTTAAAGCTTTCTCAATTTGAAGAGAT 5034
 QY 961 agtttcaagcagagatggttttagcagtaagatatttgagaaagaatgtaactcagat 1020
 Db 5033 GGTTTCAAGAAAAGACTTTCANAGTAATGTTAGAGAACTGTGCAATTTACCAAT 4974
 QY 1021 ggaacttaacttaactgaatgtaattctccagctggttatagtatccagagacaaac 1080
 Db 4973 GGGACTATATCCTTAACAGAAACATCATCTCAGATGATTAATAATTGCGGACCCGATT 4914
 QY 1081 actttaagttgaagctgagcaagtgatactat---tatttgtaaaaacagattgaa 1137
 Db 4913 AAGTTTAGAGTAGAATTAATAAAGTATTTATCTCCAAAAGAGTGGTCTCAAGTGGAA 4854
 QY 1138 aatcccaataaagagatagtagagccttactcagtagaagaataatgattt---gaa 1194
 Db 4853 AATCCAAACAAAGAGTAGACAGAGCCATCTCACTGAGAGCGTAAATGATTTATGGAT 4794
 QY 1195 gaatttagcgtttaaactacaacaaactatgcaaatlttaltatgcaaaaataaataat 1254
 Db 4793 GAAGAGTACTCTCGGGTTTATTCTCAGACGAAATTTCTATTACGGTACAAATAAGAT 4734
 QY 1255 ggaagttcaacaggtgtctattgctttaaagcagatcctaaatccacagactctgaa 1314
 Db 4733 AAAAGTTCACAAAGTGTCTACTGCTCAATGCTGATTACCTCACCACTGACTCATAT 4674
 QY 1315 gatgttggaagaacagatgactccaga---cttacaacaggaagaatgaataactcaat 1371
 Db 4673 GATAGTGTGAGACTATTAATCCAGATATCTAGTACGATGAAGAGATCAATACACAT 4614
 QY 1372 atgcaagctgctgactccttataatcttgaacaaagagatagacagactcgtgaact 1431
 Db 4613 ACGGAGAGTAGTACTGTTTAATATGCGCTAGACCGAGATACAAATCCAGAAAGAC 4554
 QY 1432 tctctaaacatacaaaaagaatattgagaaggttaccaggaagaagaagcaagctatt 1491
 Db 4553 TTTCTTAAGACATTTAAAGAAATGTAATGAAAGGCTAACAGAAAGG-----TGAT 4500
 QY 1492 gagtatagtgcttaactgagacaaatgctgctgactactcagttagcaatatattat 1551
 Db 4499 AGCTATTAATGATTAACAGAAACACAGTTTCGCGGCTACTCGCTTGCTATCTATTAT 4440
 QY 1552 ttcacgtatgctgtaattagata-----agataaactaaagaactatcat 1599
 Db 4439 TTTACAGACAGTGTGCTTAATAAAACCTTAAAACTATATACATGGAAGGTTACCAT 4380
 QY 1600 ggttttgagacatgaatgtagtactttagcagttgctaataacccctgtgtaagact 1659
 Db 4379 GGATTTAAATCTATGATGAAAGAAACCTTACGTCACAAAGAAATTAATTAATTAATGCT 4320
 QY 1660 caagatgataactccacacagctgactgactgatttcttcttccgaataacaataaa 1719
 Db 4319 CAAAATGGCAGTGGCCCTCAACTAACAATCTTATTTCTTGCTATACCTAATTAATAGCAA 4260
 QY 1720 tatcaactcttatttggaactcagtgagcatccagaagaatattgattgatatattcgatg 1779
 Db 4259 GACCAATCTCTTATTGGAGACGAATGCCATCCAGATGATTTGGTTGACGCTATTCGATG 4200
 QY 1780 gaagataaaa---aagaagtatacctgtaactcataatattcaacattgagaaacaggtg 1836
 Db 4199 GAAGATAAAAAGCAAGAGATTATTCAGTACTCACAGTTTGGACAGTGCAAAAAACAGTA 4140

QY 1837 actggttagctggtgacagaaactaaagattccatttgaattgaattgaataaataat 1896
 Db 4139 GTCCGTAGATTGGAGATTAATAACTTAAGCGCTCCAAATTTGACTTGATGAAGATTA 4080
 QY 1897 aagcAagaattgcttctcaaacgtttaaacagataaaacaaactcgaattaaagat 1956
 Db 4079 ACTGCACAGCCTATTGTTAACACTCTTAATAACTTAATATCAAGATTATTAAGTAAAGAT 4020
 QY 1957 ggtaaagcaacataatttaaaacatggggaaagtttaacactcgaagtttaaccagaa 2016
 Db 4019 GGGAAATATTCATTTAATCTTAAGACATGTCACACCTAAGAAATGAAGATTAACGACG 3960
 QY 2017 ggtattcttaactctgcaagaacagattctcgaagccttaaggttaagtttaacag 2076
 Db 3959 GGATATTTCTTATCTCGAAGAGCGCTGAAGCTAAGATTATATAGTAAACCTTGATTAAC 3900
 QY 2077 caagaagtagcaaatgctacagtttcaaaaacaggaataaacaagtgatgagacactgct 2136
 Db 3899 AAAGTTAGTCAAGACGCGAGTCAAGTAGTAGATTAACAGAAAGACAAAAGATCACT 3840
 QY 2137 ttgaaataataaagagcctggtgttccctacagagagttatcaaaagatcaatgctat 2196
 Db 3839 TTTGAAAACCGAAAGATCTTTGCCACCAACTGTTGACAAACAGATGGGGCTATCTAT 3780
 QY 2197 ctgaacttgatagtagcgtgtagatcagtttgagagctcggagaaatccacagataag 2256
 Db 3779 CTGTGGTTGTTATTAATCTGTTCCACTTGCGGTTATGTTGGCTATTTGGTCGTAAGGG 3720
 QY 2257 ataagaaacatgact 2272
 Db 3719 TTAATAAATGACTAAT 3704

RESULT 2
 STRPTNF
 LOCUS
 DEFINITION
 complete cds.
 ACCESSION
 L10919
 VERSION
 L10919.1 GI:425479
 KEYWORDS
 adhesin; fibronectin-binding protein; protein F; prf gene.
 SOURCE
 Streptococcus pyogenes (strain JRS75) DNA.
 ORGANISM
 Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae; Streptococcus.
 REFERENCE
 1 (bases 1 to 2845)
 Sela,S., Aviv,A., Burshten,I., Tovl,A., Caparon,M.G. and Hanski,F.
 TITLE
 Protein F: An adhesin of Streptococcus pyogenes binds fibronectin
 via two distinct domains
 JOURNAL
 Mol. Microbiol. 10, 1049-1055 (1993)
 MEDLINE
 95020565
 FEATURES
 source
 location/Qualifiers
 1..2845
 /organism="Streptococcus pyogenes"
 /strain="JRS75"
 /db_xref="taxon:1314"
 425..2404
 /gene="prf"
 425..2404
 /gene="prf"
 /codon_start=1
 /transl_table=11
 /function="binds fibronectin"
 /product="protein F"
 /protein_id="AA26964.1"
 /db_xref="GI:425480"
 /translation="MEVGEKMNKIFLNKEASFLAHTKRRRAVTLVGFPMLLAC
 AGAIGOVVYADDEKVPESHSSPNPEFPGYDAYGKEPGYINTRYHDLVNLNG
 SRSYGVYCNFIOISNYPSSOKSFKNPFKTEGSKSPVDAHTTKGKELEBRLLS
 LYNAYRNDANGYKKGLEHNLATVIOYAWHYHSDNIOYCEFTLMESEAKRGKSRSOY
 TLWRDLAKLIDNLEATAVNKIPSGIRLTFESDENADONLSAETVEDDPPKPEET
 SEHNPTPEIDGPIPEDPKRHPDNLPEPLTPVMLDGEVEPVSLEPALPPLME

LDGQEVPEKPSIDLPTEVPRYEFNNKQSLPAGESGETEYITFVGNQONPVDIKLL
 PNRTGSGNNVETEDTKKEPVLKMGQSESEVEFTKPDOTGSGQTTPOVEVEDTKKEP
 LMGQSESEVEFTKPDOTGSGQTTPOVEVEDTKKEPVLKMGQSESEVEFTKPDOTGSGQ
 OTTPOVEVEDTKKEPVLKMGQSESEVEFTKPDOTGSGQTTPOVEVEDTKKEPVLKMGQ
 SESEVEFTKPDOTGSGQSESEVEFTKPDOTGSGQTTPOVEVEDTKKEPVLKMGQ
 ATGDINENLAFILILISLVSIFSLKKNQSNKKV"

repeat_unit 1337..1457
 /gene="ptf"
 /note="repeat domain 1"
 repeat_unit 1631..2170
 /gene="ptf"
 /note="repeat domain 2"

BASE COUNT 977 a 472 c 616 g 780 t
 ORIGIN

Query Match 9.5%; Score 215.8; DB 3; Length 2845;
 Best Local Similarity 57.2%; Pred. No. 6,7e-26;
 Matches 466; Conservative 0; Mismatches 322; Indels 27; Gaps 3;

QY 51 ggtatgaataaaactcaaaagattactgtcacttagtgagagctctttaatgat 110
 DB 487 GGCAACACCAAAAGGCAATTTGCTGCTCACTTTAGTGGAGCTTTTATGCT 546

QY 111 ctgcgcttgtaactccatgctgtgtcgaagctgttttgtagtagatcctc 170
 DB 547 TTTGGCATGCGCGGCTGCTATCGTTTGGTCAAGTACCTATCTGCCGTGAAGAC 606

QY 171 gacgcacaacgaataaactcaagatcgaagatcgaagatcgaagatcgaagatc 230
 DB 607 TGTGCCAGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 660

QY 231 ttatgtaagagagcattccatataataaagattagagtagacagattaaaggttaa 290
 DB 661 TTAAGTAAAGCAATATCTGATTAATTAATTAATTAATTAATTAATTAATTAATTA 720

QY 291 cttagaaggaagtagaagtagaagtagaagtagaagtagaagtagaagtagaagtag 350
 DB 721 TTTAAAGGAGAGTGGAGCTGCAAGTTTATGTTTAAATTAATTAATTAATTAATTA 780

QY 351 cgaatcagatagtagtctaaagattgtaaaacatgatgatcctcaaaaatt 410
 DB 781 CCAAAAATATGTTTCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 840

QY 411 tgaagatagtagagtagccttagaattacggagatgagcattcaagaagtagagc 470
 DB 841 TGTAGATTATGCTACACTATAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAA 900

QY 471 tgtatgataatgagacatccaaataatgccaatggtatataatgaaagcttggaac 530
 DB 901 TTTACTGATTAATGCTATGCTATGCTATGCTATGCTATGCTATGCTATGCTATGCT 960

QY 531 gaatcagatagtagtcaaaagagcgtatggtactatctgataatgctcctattc 590
 DB 961 AAAGTATTAATGCTATGCTATGCTATGCTATGCTATGCTATGCTATGCTATGCT 1020

QY 591 taatcagatagtagtcaaaagagcgtatggtactatctgataatgctcctattc 650
 DB 1021 ATTTGAGACATTAATGAGAGTCTGAGGCGAAAGAGGTTA--GATTAATGAGTCA 1077

QY 651 atcttgatcgtcaagcttggaagcaactgattgattcgaatttgcaactaaatgccc 710
 DB 1078 TACTTAAATGCTATGCTATGCTATGCTATGCTATGCTATGCTATGCTATGCTATG 1137

QY 711 aaacagatcgtgagtagtattcagctaaglatcttgagctcaggaaggaagataa 770
 DB 1138 TAAATAAATGCTATGCTATGCTATGCTATGCTATGCTATGCTATGCTATGCTATG 1181

QY 771 atataaagaagatacaaaaactcttgagtagtgtgtgttgtagtctcaataaccaaac 830
 DB 1182 --AAAAAGAGCTTAAATGCTATGCTATGCTATGCTATGCTATGCTATGCTATG 1239

QY 831 tccagagagaccacacatgctcctcaaatcaacc 865

DB 1240 ACCTGTGAGACGTCAGACATTAATCTTAATAATC 1274

RESULT 3
 SPSFB
 LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM

SPSFB 2263 bp DNA BCT 13-JUL-1995
 S.pyogenes Sfb gene for fibronectin-binding protein (partial).
 X67947 S42389
 X67947.1 GI:511149
 fibronectin-binding protein; sfb gene.
 Streptococcus pyogenes.
 Streptococcus pyogenes.
 Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
 Streptococcus.
 1 (bases 1 to 2263)
 Talay,S.R.
 Direct Submission
 Submitted (31-JUL-1992) S.R. Talay, GBF, Mascheroder Weg 1, 3300
 Braunschweig, FRG

REMARK
 REFERENCE
 AUTHORS

TITLE
 JOURNAL
 MEDLINE
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL

1 (bases 1 to 2263)
 Talay,S.R., Valentín-Weigand,P., Timmis,K.N. and Chhatwal,G.S.
 Infect. Immun. 60 (9), 3837-3844 (1992)
 3 (bases 1 to 2263)
 Talay,S.R., Valentín-Weigand,P., Timmis,K.N. and Chhatwal,G.S.
 Domain structure and conserved epitopes of Sfb protein, the
 fibronectin-binding adhesin of Streptococcus pyogenes
 Mol. Microbiol. 13 (3), 531-539 (1994)
 4 (bases 1 to 2263)
 Talay,S.R.
 Direct Submission
 Submitted (13-JUL-1994) S.R. Talay, GBF, Mascheroder Weg 1, 3300
 Braunschweig, FRG
 On Jul 17, 1994 this sequence version replaced g1:47433.
 location/Qualifiers
 1..2263
 /organism="Streptococcus pyogenes"
 /strain="DSM 2071"
 /db_xref="taxon:1314"
 124..270
 /gene="Sfb"
 124..2040
 /gene="Sfb"
 124..2040
 /gene="Sfb"
 /transl_table=1
 /product="fibronectin-binding protein"
 /protein_id="CAA48133.1"
 /db_xref="GI:511150"
 /db_xref="SPTREMBL:O01924"
 /translation="MNNKFLNKEAGFLAHTKKRRFAVTLGVFPMLLASAGATGCG
 OVAYADDEKTVPHRVSONPEFPWYGYDFKGYPTRYHNLNLNGSKSTVOACFNLR
 FEKKSGSEFPNNYKRMDSSEFVYADVPDRDNSSRYIDVLEKNTLRVLYNGYP
 NNGNGIMEGIEATLITONAVMYSDNSIFPNDNFTEAKDKDNLKREOLSLRV
 ALKKLLDPKISESLKRPVSTFRNLNFEESODKLYKMLSAEYEPKPRKGETPEKQ
 KPELDGTPLEGPORPNSLEPTLPVMDGQVEVPSSELEPALPLMPELDQOE
 VPEPSPSELEPALPLMPELDQOEVEKSVLDLPIDPRYEFNNKQDSQSEVLEKSGTE
 YTEVYGNQONPVIDIKLPLENTEGFGSNMVEEDTKPEVLMGQSESEVEFTKPDOTG
 MSGOTPOVEVEDTKKEPVLKMGQSESEVEFTKPDOTGMSGOTPOVEVEDTKKEPVLK
 GQSESEVEFTKPDOTGMSGOTPOVEVEDTKKEPVLKMGQSESEVEFTKPDOTGMSG
 EYTVIEDTRPKIVFHFNDNNEPVEENREKPTKNTITPILPATGDINENLAFILILIS
 VLPFISLKKRQKQ"

BASE COUNT 790 a 383 c 477 g 613 t
 ORIGIN


```

Db 717 AATGATATCCAAATGATGCAATGAGATTATGAAAGGATTGGACCGTACAAATGCTATT 776
QY 541 aaggttcaacaagaagcggtatgtaactatctcgataatctccttaattccatccagat 600
Db 777 TTAGTAACCTCAAACTCGGATTGGTATTATTCGATNA---CTCAAGATTATAGTGAT 833
QY 601 gaaagtttaaaagagagctagaagaacttgtagtaactcttccttaattctttagat 660
Db 834 GAACAGTCGGAAGCTGTGATTAAGAGATGAGAAATTAAGCAAAATCAAGTAAGATTATAG 893
QY 661 cgtcaagctttgaagcaactgattgattccgaatttggcacaactaaatgcacaagaagt 720
Db 894 CGGATGCTCTGTGAAATAATATATATACACGATTGTAGAAAAACAGCAAAAAATCACTTG 953
QY 721 cggatgatttcaagcttaagtatttttgagctcgaagcagaaggagataaat 774
Db 954 CCGAGATTAATTAAGTTAAATCTTTTATTTCAAAAGATACCTGATCAAAAT 1007

RESULT 5
LOCUS AF009908 2801 bp DNA BCT 13-FEB-1998
DEFINITION Streptococcus pyogenes fibronectin/fibrinogen binding protein F
ACCESSION AF009908
VERSION AF009908.1 GI:2271466
KEYWORDS Streptococcus pyogenes.
SOURCE Streptococcus pyogenes.
ORGANISM Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
Streptococcus.
REFERENCE 1 (bases 1 to 2801)
AUTHORS Katerov,V., Andreev,A., Schalen,C. and Totolian,A.A.
TITLE Protein F, a fibronectin-binding protein of Streptococcus pyogenes,
also binds human fibrinogen: isolation of the protein and mapping
of the binding region
JOURNAL Microbiology 144 (Pt 1), 119-126 (1998)
MEDLINE 98129085
REFERENCE 2 (bases 1 to 2801)
AUTHORS Katerov,V.E., Andreev,A.S., Schalen,C. and Totolian,A.A.
TITLE Direct Submission
JOURNAL Submitted (20-JUN-1997) Med. Microbiol, Lund University, Solvegatan
23, Lund 22362, Sweden
FEATURES
source location/Qualifiers
1. 2801
/organism="Streptococcus pyogenes"
/strain="Efi1949"
/serotype="M15"
/db_xref="taxon:1314"
450..2507
/genes="prtF15"
450..2507
/genes="prtF15"
/codon_start=1
/transl_table=11
/product="fibronectin/fibrinogen binding protein F"
/protein_id="AAC38155.1"
/db_xref="GI:2271467"
/translacion="MFLRGEKMNKKIFLNKEASFLAHTRRKRFAVTLVGFEMLFAC
ASISIGGVAAVADEKTEPNRINSNPEFMYGDSGALLRHNLKVNGSKEYOAYC
FLNKREPKKESSSNMYKKLDGSTETPKYKAENRFSGEELRHILKLVINGYPS
NEIMKIDPLNAILYTONAIIYYSASAPINDINNPTSEANDLINTPOOLITRMLR
KLISDBENLVKQPSNFKLSTFESSDKSTONLSAIVVDPPKPDGTSBNKTPLE
DGTPIEDPKRDESSGAPLPLMPELDGEVEVSESLERLPLVPELDGEVE
VSESLERLPLMPELDGEVEVSESLERLPLMPELDGEVEVSESLERLPL
PLMPELDGEVEVSESLERLPLMPELDGEVEVSESLERLPLMPELDGEVEV
DIDKRLPNETFGSNVETETKREKGVLMGQSESEVETKQDQAGSGGTTQVEE
THEPEVLMGQSESEVETKQDQAGSGGTTQVEE
QAGMSGQTAPOVETEDTKPEVLMGQSESEVETKQDQAGSGGTTQVEE
VFHFDNNEKVEHREKPTKNITPILPATGDIENLAFGLILSVLSIFSLKNNK
NRY"
BASE COUNT 954 a 489 c 597 g 761 t
ORIGIN

```

```

Query Match 6.68; Score 150.2; DB 1; Length 2801;
Best Local Similarity 54.68; Pred. No. 2.8e-15;
Matches 445; Conservative 0; Mismatches 313; Indels 57; Gaps 5;

QY 51 ggtattaagtaaaacacaaacaaagattactgacttgtagtggaagcttttaattat 110
Db 512 GGCACACACAAAGAAAGAAAGGCGATTGGCTGCACCTTAGTGAGAGCTCTTTTATGCT 571
QY 111 ctgcgcttggtaactccaatggttggctgaagactgcttggctttagtaaatcc 170
Db 572 TTTTGATCGCGGGGTTCTATCGTTGGTCAAGTACCTATGCGCGGATGAGAAAG 631
QY 171 gaagccaagcaataaataatccagattcaagtcggaatacagatgtaagataatc 230
Db 632 TGAGCCGAATCGTATTA-----ATTCAAAATCCTGAGTTCTTGATGCTATGATTC 685
QY 231 ttatgtaagaggcaatccatattataacagctttagagtagcaacagattaaaggttaa 290
Db 686 GTATAGCGGAAGATTGTTAAGATATCACAACTGAAAGTNA----- 726
QY 291 cttagaaggaagtagaagatatacaagttatcttgaatttaagaagaacattccctc 350
Db 727 -----ATGGAAGTATAGAGATATCAAGCGTATGTTTAACTGAAAGGTTGACGCCAA 781
QY 351 cggatcagatagtagtgttaaaagtgttaaaaaaacatgatgtaatctcaaaaat 410
Db 782 AAAAGAAAGAAAGTAGACAGTCTTAATGCTATTAATAAATTAAGAGGAGTACAGAACTTT 841
QY 411 tgaagattatgcgatgagacccatgaatttaaggagatgagctaaacagagatgagc 470
Db 842 TAAATAATFACGCTGAAGAACCTTGTTTAACTGAGAGGAAATTAAAGACGACATATTTGAA 901
QY 471 tgtatgataataagagacatccacaataatgccaatggtatattatgaaaggtctggaac 530
Db 902 GGTTCCTCTACAAATGATGATCC---TATATATATTAAGAAATTAAGAGGATATACCTCT 958
QY 531 gaattctatacagaatcacacaagaagcgtagtactatcttgataatgctctatttc 590
Db 959 CAACGCTATCTGTGAACACAAATAATGCCATATGCTATTAATCTCGAATAGTCCACCGATTAA 1018
QY 591 taatccagatgaagaatttaagaaggagatcagaagaacttgtagtaactctccta 650
Db 1019 TGAATTAATTAATACCTCTTCTACTAGTGAAGCAATGATCTTAATATACCTACAGACGCT 1078
QY 651 atcttgatgcgtcaagctttagaagcaactgattgattccgaatttggcacaactaaatgc 710
Db 1079 AAGCTTAATCGGAGAGCTGTGAGAAACATAATTTCTCTGAT-----GAAATTTTGGT 1132
QY 711 aaaaagaattcggatgatttcaagctaaatatttgaagctcgaagcagaaggagataa 770
Db 1133 TAAACAAAGTCCATCGAACCTTTAACTTAATTTTGAATCT----- 1175
QY 771 atataataaagatatacaaaatctttagtggttgaagcttactactaaaccacaac 830
Db 1176 -TCGTAAATCTCTACCAAAATCTTTTAAGTGTGTAATATGTAATGATGATCTCTCTTA 1234
QY 831 tccaagagaccacacaaatgcctccaatcaactc 865
Db 1235 ACCTGGATGATACGTCAAGAACATATATCTTAATAACTG 1269

RESULT 6
LOCUS AF009914 321 bp DNA BCT 13-FEB-1998
DEFINITION Streptococcus pyogenes fibronectin binding protein F gene, partial
ACCESSION AF009914
VERSION AF009914.1 GI:2267181
KEYWORDS Streptococcus pyogenes.
SOURCE Streptococcus pyogenes.
ORGANISM Streptococcus pyogenes

```


JOURNAL		Submitted (23-JUN-1997) Medical Microbiology, Lund University, Solvegaten 23, Lund 22362, Sweden	
FEATURES		Location/Qualifiers	
source		1. 348 /organism="Streptococcus pyogenes" /strain="M18 strain SS-36" /db_xref="taxon:1314"	
CDS		<1..348 /codon_start=2 /transl_table=1 /product="fibronectin binding protein F" /protein_id="AAC38167.1" /db_xref="GI:2267194" /translation="FGGVATVAADKTYVPNEKSPDDYPWGYDAFLFYHDLNNV LESTPQVYVCFMLVROEPSKVNCFRRNWKVDGDAVFKKYANPRVIDRRTN LNIYNGYSRVOMG"	
BASE COUNT	115 a	45 c	78 g 110 t
ORIGIN			
Query Match	3.2%;	Score 72;	DB 1; Length 348;
Best Local Similarity	55.6%;	Pred. No. 0.017;	
Matches 138;	Conservative 0;	Mismatches 110;	Indels 0; Gaps 0;
Oy	245	atccatattataaacagtttagagtagcacacagatttaagggttaactagaagaagta	304
Db	87	ATGCTTAACACGCGCGCTTTTAAGATATCATGATTTAAATTTGGAAAGGTACGA	146
Oy	305	gaagttacagattatttgcttatttaagaagaagcattccctcgcgatcagtagta	364
Db	147	CGCCTTACCAAGTTTATTGCTTTAATTAGTAAGACAGCAACCGATTAAGTGCTT	206
Oy	365	gtgttaaaaagttggtatataaaaacatgtagtaatctcacaaaatttgaagatagca	424
Db	207	TTAGAAAAAATTTGGTTCAAAAAAGTGAATGCGATGAATGCTGTGTTAAAAAATATGCG	266
Oy	425	tgagccctagaattacaggagagatgaagctaaatcagaagttacgagcttcatgtaaty	484
Db	267	CTAATCCATGAGTAATGATGCGAGATTTAGAACGACATTTTAATGTAAATTTCACAG	326
Oy	485	gacatcca	492
Db	327	GATATTC	334
RESULT 9			
LOCUS	AC007926		
DEFINITION	AC007926 149789 bp DNA	HTG	12-FEB-2001
ACCESSION	Trypanosoma brucei chromosome II clone RPC193-3H15, *** SEQUENCING		
VERSION	AC007926		
KEYWORDS	AC007926.7 GI:12746529		
SOURCE	HTG; HTGS_PHASE2.		
ORGANISM	Trypanosoma brucei.		
REFERENCE	Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.		
AUTHORS	1 (bases 1 to 149789) El-Sayed, N., Zhao, H., Mason, T., Wanless, D., Fujii, C., Barnstead, M., Bowman, C., Peterson, J., Khalak, H., Ullu, E., Melville, S., Donelson, J., White, O., Fraser, C. and Adams, M.		
TITLE	Trypanosoma brucei GUMat10.1 RPC193-3H15 BAC genomic sequence		
JOURNAL	Unpublished		
COMMENT	2 (bases 1 to 149789) El-Sayed, N.M., Khalak, H. and Adams, M.D. Direct Submission Submitted (28-JUN-1999) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA On Feb 12, 2001 this sequence version replaced gi:6707792. * NOTE: This is a 'working draft' sequence. It currently * consists of 1 contigs. Gaps between the contigs * are represented as runs of N. The order of the pieces * is believed to be correct as given, however the sizes		

[illegible]

*	1	133919	133951:	contig of 133918 bp in length
*		133952	143677:	contig of 9626 bp in length
*		143678	143610:	gap of unknown length
*		143681	147695:	contig of 3885 bp in length
*		147696	147828:	gap of unknown length
*		147829	151430:	contig of 3602 bp in length
*		151431	151562:	gap of unknown length
*		151563	154701:	contig of 3139 bp in length
*		154702	154833:	gap of unknown length
*		154834	157774:	contig of 2941 bp in length
*		157775	157906:	gap of unknown length
*		157907	160810:	contig of 2904 bp in length
*		160811	160942:	gap of unknown length
*		160943	163756:	contig of 2814 bp in length
*		163757	163888:	gap of unknown length
*		163889	166433:	contig of 2545 bp in length
*		166434	166565:	gap of unknown length
*		166566	166996:	contig of 2431 bp in length
*		166997	169126:	gap of unknown length
*		169129	171451:	contig of 3232 bp in length
*		171452	171583:	gap of unknown length
*		171584	173757:	contig of 2192 bp in length
*		173758	173907:	gap of unknown length
*		173908	175981:	contig of 2074 bp in length
*		175982	176113:	gap of unknown length
*		176114	178179:	contig of 2066 bp in length
*		178180	178311:	gap of unknown length
*		178312	180315:	contig of 2004 bp in length
*		180316	180447:	gap of unknown length
*		180448	182441:	contig of 1994 bp in length
*		182442	182573:	gap of unknown length
*		182574	184483:	contig of 1910 bp in length
*		184484	184615:	gap of unknown length
*		184616	186473:	contig of 1858 bp in length
*		186474	186605:	gap of unknown length
*		186606	188459:	contig of 1854 bp in length
*		188460	188591:	gap of unknown length
*		188592	190422:	contig of 1831 bp in length
*		190423	190554:	gap of unknown length
*		190555	192291:	contig of 1737 bp in length
*		192292	192423:	gap of unknown length
*		192424	194078:	contig of 1655 bp in length
*		194079	194210:	gap of unknown length
*		194211	195801:	contig of 1591 bp in length
*		195802	195933:	gap of unknown length
*		195934	195969:	contig of 36 bp in length.

```

/organism="Oryza sativa"
/db_xref="taxon:4530"
/clone="OSJNna00b05C16"
BASE COUNT      58424  a  38750  c  34838  g  53551  t  10406  others
BRIGIN

```

Oy	774	taataaaagataccaaatcttltgagtgggttgattcctaactaacccaaccc	833
Db	148741	TNNNAANNNNNNNNAANNNNATTAATTTTTTTTTANNNNNNTNNAANTNNANNNNNAAT	1488000
Oy	834	aggagaccacaacatgcctccaatcaacctcaacgaagcttagtacttatagaagta	893

Db 148801 ANNTANAAAAATTAATTAATAAANAANAANAANAANAATNNAAAAATTATANNNNNA 148860
Qy 894 tgcataagtgatctactcctaattcgctggaagcagcactacagctgacagggcataa 953
Db 148861 NANTTAATTAATNNATNNATTAANNTATTAANNAANNNNNNNNNAAAAAATTAATTTNNN 148920
Qy 954 cgtgaatgcttccaagcgcagagtgctttagcagtaatgatacttggagaagaatcgaact 1013
Db 148921 NNANAAAAATNNATTAANNAANNNNNNNATNNNNNNATNAANAANAANNTAAANNAANNNNNNA 148980
Qy 1014 atcgaatggaactatactcttaactgaaatcgaattcccgccggttatagcttcgcaga 1073
Db 148981 TNANNNNNNNNNNNNNAANNTAANAANNNNTNTNATTTTATATATATATATANATAT 149040
Qy 1074 gccactcactttagagctggaagctgcgaagctgatacctatttgatggaacaacagat 1133
Db 149041 ATTAAANNNTTAANNTPAANNNAANAANAANNNNNNNNNNAANAANAANNAATTAANAATNNAN 149100
Qy 1134 tgaataatcccaataagaagatagtagagccttactcagtagaagcatalatgatttga 1193
Db 149101 NAAAAAATNAAAAAANAANAANAATTTTNTNNNNNAANAANNTAAANANATVANNTNNA 149160
Qy 1194 agaatttcgcttactacacacaaactatgcataatttatatgcaaaaaataaaa 1253
Db 149161 ANTTTAAATTAATTATNNAAAAAANAANAANAANAANAATTTAAATTTAAAAAANNNANNTT 149220
Qy 1254 tggaaagtcacaggtctgctctatgctttaaagtcagactaaactccacgcagctcga 1313
Db 149221 TTTAAA---AAATANTTTTAAANNTTTTAAAAAANAANNTTAAAAANNNNNNAANAANAAN 149277
Qy 1314 agatgctgggaacaacatgactccagacttacacacagagaagtaaatatacatcat 1373
Db 149278 NNANNTTNNNNNNNAANNTTAAATTAANNTAAANNNNNNANNTTTNNNNNNNNNN 149337
Qy 1374 tgcaggtcgtgacctcttaatactatctgigaacccaagataccgatcccgacctt 1433
Db 149338 NNANATNTNTTANNNNNNNNNNNTTAAANNNNAANAANNNNNNANANGNNAATTAANNNNA 149397
Qy 1434 cttaaacatactcaaaaaagtaacttggagaagctacagggaagaagcagactatga 1493
Db 149398 AAANNAANNTTAANNNTAANNAANNNNNNAANAANAANAANNAANNNNNNAANNNAGN 149457
Qy 1494 gtaatgctgaactcagacacaaatctgcgtgcgcactcagctagcaatatattt 1553
Db 149458 NNNTTAAATTTAANNNNNNNNNNNANNTTANNNNAANNNNTTAAANNNANNNNNNANNTTA 149517
Qy 1554 cactgatactgctgaatagataagataaaactaaactacatcagtcttggagacat 1613
Db 149518 NNNNNAANNNNNNNNNNTTNNNAANAANAANNTTAAAAAANNAANNAATTTTAAAAAN 149577
Qy 1614 gaatgatactacttagcagctgcttaaaatccctgctgaatagcctcagaatgatactc 1673
Db 149578 NNNNNAATTAATAAANNNNNNNNNNNNAANAATTTTANNNNNTTANNTNNANNNNNNAANNN 149637
Qy 1674 tccaacagctaacctgacctgattcttattccgaatacacaataaatacaactcttat 1733
Db 149638 NTAA-----AAAANNNANNNNNNNNTTANNTNNNNNTTNCAAAAAANAANNNNNNANNTT 149692
Qy 1734 tggaaactcagtcgacatccagaagattagctgatatattcgtatggaagataaaaaaga 1793
Db 149693 NNAAAAATTTANNNNNAANAANAANNTNNNAANNNNAANNTTAAANNNNAANAATTTAAAAANNNNA 149752
Qy 1794 agcttaccctgaactcataatctaacaatcggagaagaacggctgacgtcttgcgtgga 1853
Db 149753 AAANNTNNNAANAANAANAANAANNAANNTTAAATNNAAANNTTTNNNTNNAANAANAAN 149812
Qy 1854 cagaactaagatctccattctggaatcgtgaataaataaataaagaacgaactgcttc 1913
Db 149813 AAAANNNAAAAANNTTNNNNNNNNAAATTTNAAAAAANAANNNNNNNNTNNNNNAANNTTAAA 149872
Qy 1914 tcaaatctttaaacagataaacaacatcgaatttaaagatgataagcgaacctaa 1973

